Mental Health Prediction Among Medical Students Using Machine Learning

By
Nosrat Jahan Tithi
212-15-4105

Kazi Zakiul Haque
212-15-4101

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by
Md. Sazzadur Ahmed

Assistant Professor
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by
Mr. Shah Md Tanvir Siddiquee

Assistant Professor
Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

May 14, 2025

APPROVAL

This Project titled **"Mental Health Prediction among Medical Students Using Machine Learning "**, submitted by Name, ID No: **212-15-4105 and 212-15-4101** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **14 May, 2025**.

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mohammad Monirul Islam **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Ms. Tasfia Anika Bushra **Internal Examiner**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Name of the Supervisor, Supervisor's Designation**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Md. Sazzadur Ahmed
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by:

Mr. Shah Md Tanvir Siddique
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

Tithi

Nosrat Jahan Tithi
Student ID: 212-15-4105
Department of Computer Science and EngineeringDaffodil International University

Kazi Zakiul Haque
Student ID: 212-15-4101
Department of Computer Science and

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Md. Sazzadur Ahmed, Assistant Professor,** Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Machine Learning** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Mental health is a critical aspect of well-being, influencing individuals' ability to function in daily life and cope with stress. Medical students struggle with depression alongside their demanding academic schedule and long workdays since these circumstances harm their academic results and general health condition. The objective of this research considers the application of machine learning algorithms to detect depression severity levels in medical students. Data contains a total of fourteen features with demographic information from age and year of study and psychological indicators such as interest scores compared to pleasure levels in addition to ratings regarding fatigue and sleep problems as well as changes in appetite and concentration abilities and PHQ-9 scoring scales. The depression severity makes up the target variable which divides into four groups: Severe Depression, Moderate Depression, Mild Depression and Moderately Severe Depression. The assessment of this task incorporated numerous machine learning models which consisted of Gaussian Naive Bayes, Random Forest Classifier, AdaBoost Classifier, Logistic Regression and Support Vector Classifier (SVC). Random Forest Classifier delivered the highest accuracy of 99.04% while Logistic Regression reached an accuracy of 98.80% yet Gaussian Naive Bayes obtained 97.36% accuracy. The accuracy of Support Vector Classifier at 95.44% was lower than the other models which included AdaBoost and its weakest performance of 79.41%. Predictions of depression severity among medical students demonstrate optimal performance when using Random Forest as the model selection because of its strong predictive capabilities. The study reveals crucial roles which machine learning serves mental health prediction alongside giving researchers a way to recognize at-risk students for proper early interventions and individualized treatments. Additional optimization and more targeted feature engineering techniques seem necessary to raise AdaBoost's results since there exists room for increased model accuracy. Researchers show that artificial intelligence performs effectively for mental health detection because it helps identify different depression severity levels within medical students who face elevated stress and health challenges.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The introductory section of this report presents project rationale through examination of its motivation alongside its objectives together with methodological approach description. The initial part outlines future report results while presenting its organizational framework.

1.1 Introduction

The worldwide recognition of mental health problems particularly affects medical students who experience high levels of academic and emotional strain in their educational environment. Students enrolled in medical programs face mental health issues including depression alongside anxiety because they must handle heavy coursework requirements and extended study periods while viewing patient cases together with maintaining high performance targets. The widespread mental disorder depression proves very problematic for medical students because it leads to severe consequences that damage both their grades and physical condition and their entire lifestyle health. Physical and mental health outcomes of students can benefit greatly from early identification of depression levels because it allows immediate access to appropriate help.

The research employs machine learning methods to forecast medical student depression severity by analyzing their psychological characteristics together with demographic data. The dataset employed in this research contains 14 elements including demographic information combined with psychological assessment factors including interest and pleasure and fatigue as well as sleep issues and concentration and the standard PHQ-9 scoring method. The dataset uses depression severity as the target variable which contains four distinct categories including "Moderately Severe Depression" as well as "Moderate Depression" and "Mild Depression" and "Severe Depression."

Machine learning algorithms implemented for depression categorization of students included Gaussian Naive Bayes, Random Forest Classifier and AdaBoost Classifier, Logistic Regression as well as Support Vector Classifier (SVC). Among the applied models the Random Forest Classifier demonstrated the best performance while Logistic Regression and Gaussian Naive Bayes took second and third places based on accuracy measures. The research findings show that machine learning technology effectively evaluates depression severity which makes it possible to develop specific mental health intervention approaches for college students.

1.2 Motivation

The main reason for this research stems from increasing concern regarding medical student mental wellness because these students endure heavy academic pressures during their educational journey. Medical students experience mental health issues like depression along with anxiety and burnout because of the challenging requirements in their education together with extensive study time and demanding clinical work and high-performance standards. The timely diagnosis of depression proves essential because it affects academic success and personal well-being and determines professional achievement in the long run. Detecting medical students at high risk of severe depression remains complex while medical professionals frequently fail to identify students who need immediate care. The main purpose of this study employs machine learning methods to construct a predictive model for medical students' depression severity while analyzing their psychological and demographic characteristics.

1.3 Objectives

1. The research goal is to forecast depression severity in medical students by using the machine learning algorithms under its psychological and demographic parameters.

2. It is compared the performance of the different models such as Gaussian Naive Bayes, Random Forest, AdaBoost, Logistic Regression and SVC classifying various depression levels.

3. The research singles out essential findings in the form of PHQ-9 scores, fatigue, concentration problems and lack of interest in daily activities as main predictors for correct determination of depression-type classification.

1.4 Methodology

The research methodology commences through data compilation which requires data preparation before analysis. The analysis dataset contains 14 features that cover demographic information and psychological indicator data points including sleep issues and PHQ-9 scores and fatigue measurements. The preprocessing work includes treating missing data points while encoding categories and applying normalization strategies to numerical variables. The prediction of depression severity uses Gaussian Naive Bayes together with Random Forest Classifier as well as AdaBoost Classifier and Logistic Regression and Support Vector Classifier (SVC). The depression severity classification models achieve evaluation through dataset training to determine their ability in distinguishing depression levels across four severity groups. The comparison of the machine learning models utilizes accuracy alongside precision and recall together with F1-score performance metrics. Further development of the selected best-performing model includes hyperparameter optimization and validation procedures for achieving maximum performance.

1.5 Project Outcome

This project accomplishes the task of creating an accurate machine learning model which identifies depression severity in medical students. The Random Forest Classifier performed best among different psychological and demographic features since it reached 99.04% accuracy in model prediction. The developed model serves as an effective method which can accurately categorize student depression levels between "Moderately Severe," "Moderate," "Mild" and "Severe" depression. This research proves that machine learning systems can help identify severe depression symptoms in students before they worsen so appropriate assistance can be developed for each individual case. According to the research findings the essential features of fatigue and sleep issues together with PHQ-9 scores act as primary predictors for mental health difficulties. This work advances the practical use of machine learning tools in mental health care by introducing an automated system that benefits medical student wellness operations.

1.6 Organization of the Report

The report divides its content into six main chapters. The Introduction chapter of this report presents key elements such as project background alongside objectives and methodology together with motivation for studying mental health issues among medical students. The existing research on medical student mental health forms the foundation of Chapter 2 Background while the chapter identifies unattended areas where this project stands to make an impact. The data collection methodology and system design process appear in Chapter 3 in addition to describing the project's technical implementation. The fourth chapter describes the installation protocols while appreciating model performance through testing and result analysis alongside an evaluation of model advantages versus disadvantages. Chapter 5 of this report covers standard regulations and design obstacles alongside environmental and social effects and ethical supervision of the project. Chapter 6, Conclusion, summarizes research findings before proposing project limitations together with future research recommendations while analyzing machine learning support potential for medical student mental health.

Chapter 2

Background

This chapter includes study and research material pertaining to the project while exploring comparable applications and scientific findings connected to it. The gap analysis shows the existing unexplored and unaddressed areas within the project.

2.1 Introduction

The ability to predict and classify mental health conditions in medical students requires heightened importance because students in stressful environments experience significant psychological demands. ML techniques demonstrate impressive potential to forecast university students' mental health problems including depression moderation and anxiety development and stress responses primarily among medical school students. Research analyzes how ML models can recognize mental health risks in students through the evaluation of academic performance together with behavioral and life-style indicators. Mutalib and Shafiee (2021) showed that ML models succeed at categorizing students through mental health assessments which includes stress and anxiety and proves the strength of ML algorithms for mental health assistance [1]. Baba and Bunji (2023) applied ML algorithms to health survey data which allowed them to forecast mental health problems among students while identifying academic and social element factors that affect well-being [2]. Proof of concept research by Kabir and Khanam (2021) demonstrated how health behavior measurements of sleep quality and exercise patterns enable predictions of university student mental health status [3]. Alzubaidi et al. (2023) investigated AI-based tools for early mental distress detection in medical students which led to timely interventions and support according to their research findings [4]. ML models serve as tools for predicting student mental health outcomes and supporting university-led psychological crisis management according to Ge et al. (2021) [5]. The educational success of student mental healthcare evaluation relies heavily upon machine learning analysis of stress and burnout indicators as demonstrated through Tyulepberdinova and Mansurova's (2024) research [6]. The work of Alzubaidi et al. (2024) develops predictive algorithms for medical student mental health which enhances ML technique effectiveness in identifying warning signals before intervention becomes necessary [7]. Marcon et al. (2020) demonstrated the utility of prediction models along with machine learning methods to measure suicide thoughts and stress in medical students which indicates machine learning is effective for mental health resource enhancement [8]. The integration of questionnaire data with ML by Herbert et al. (2021) has proven the ability of such models to detect psychological well-being status among university students thus validating their use in digital mental health platforms [9]. Herbert et al. (2020) developed their previous research by deploying machine learning methods to analyze how personality factors affect student mental health showing behavioral patterns alongside ML provide precise predictions of student well-being [10].

2.2 Literature Review

Table 2.1: Summary of Literature Reviewed.

Authors	Year	Domain	Accuracy (%)
Islam, M. & Uddin, M.	2020	ML in mental well-being screening	87.0%
Rathi & Mishra	2023	Depression & anxiety via ML	90.1%
Tanveer, M. I., et al.	2022	Early intervention system for med stud.	88.9%
Zhang, L. et al.	2025	Mental health assessment using LSTM	95.0%
Our Work	2025	Mental Health Prediction Among Medical Students	**99.04%**

2.2.1 Similar Applications

Healthcare professionals actively implement machine learning to detect mental health disorders early and chronic diseases while using it to foresee treatment results. Clinical researchers utilize ML models to forecast depression and anxiety levels in patients through the analysis of behavioral and social patterns alongside physical activity measurements in ways that correspond to medical student mental health assessments. The application of ML algorithms in cardiology identifies heart disease risks by processing patient histories along with lifestyle factors and oncology makes use of image processing for cancer detection during early stages. Machine learning technologies in education serve functions that move past mental health prediction by providing tools for measuring academic results and academic achievement. The identification of students likely to drop out uses ML models to analyze engagement metrics together with attendance information and academic results just like the method for medical student stress prediction and anxiety detection processes. Multiple applications show consistent use of extensive data to produce reliable predictions as well as early detection of high-risk populations before suitable interventions can be provided. Multiple industries thrive with the help of machine learning systems because this technology demonstrates strong reliability in resolving complex healthcare and educational problems.

2.2.2 Related Research

Many studies have thoroughly analyzed the application of machine learning (ML) techniques to forecast mental health conditions within academic institutions. Research has dedicated efforts to create algorithms that predict depression and anxiety with stress as well as stress-related disorders among university students. Mutalib and Shafiee (2021) showed ML algorithms can predict mental health conditions through analysis of behavioral and academic student data according to their study [1]. The usefulness of ML in mental health diagnosis is demonstrated by survey data analysis conducted by Baba and Bunji (2023) which provided more evidence of ML's effectiveness in this field [2]. The research conducted by Alzubaidi et al. (2023) showcased AI-based tools with AI-based tools that monitor academic students for initial signs of distress thereby illustrating machine learning as a vital method for immediate responses [3]. The research by Baba and Bunji together with other studies enhances evidence which establishes ML as a powerful predictive instrument for mental health assessment in educational institutions.

2.3 Gap Analysis

Table 2.2: Gap Analysis

Study	Key Findings	Gap Identified
S. Mutalib, N. S. M. Shafiee (2021)	ML models classified students into mental health categories (stress, anxiety, depression).	Lack of integration with real-time data for dynamic intervention.
A. Baba, K. Bunji (2023)	Survey data used to predict mental health problems (depression, anxiety) based on student behavior.	Limited scope in terms of variables—did not include physical health data.
M. Alzubaidi, H. Shah (2023)	AI tools identified early signs of distress among medical students.	Need for further validation of AI-based interventions in real-world educational settings.
F. Ge, D. Zhang (2021)	ML predicted stress among university students using behavioral data such as sleep patterns and activity.	Behavioral data alone may not fully capture mental health status—integration of academic data needed.
G. Tyulepberdinova, M. Mansurova (2024)	ML models predicted overall student well-being by evaluating physical, social, and mental health data.	Further research needed on how these models interact with diverse student populations and learning environments.

2.4 Summary

A survey of machine learning methodologies applying to the prediction of mental health risks for medical students formed the main content of this chapter. The study demonstrated how ML models employ various factors including academic performance and behavioral data and lifestyle choices to recognize students suffering from mental health problems including stress and anxiety plus depression symptoms. ML-based research shows it identifies student mental health conditions by categorization while also detecting initial signs of harm so intervention strategies become possible. The discuss focused on identical ML applications across healthcare and education sectors because these analytics models perform efficiently against complicated problems. Research into related subjects confirmed how effective ML proved in mental health forecasting yet researchers noted the lack of immediate data integration along with the necessity of testing AI-based intervention approaches in actual environments.

Chapter 3

Research Methodology

Requirement analysis and system design along with specification are explained in this chapter through methodology details. The project specification comprises functional and nonfunctional requirements while illustrating design diagrams in addition to providing a complete project plan that includes task allocations.

3.1 Methodology/Requirement Analysis & Design Specification

3.1.1 Overview

The research procedure for this initiative requires a series of essential actions beginning with data gathering and preparation work. The dataset consists of features such as demographic information together with psychological indicators and PHQ-9 scores that provide the basis for depression severity prediction. Prior to analysis the data needs preprocessing steps which involve treating missing values while translating categorical data and applying normalization to numerical data to create consistent data and enhance model performances.

3.1.2 Proposed Methodology/ System Design

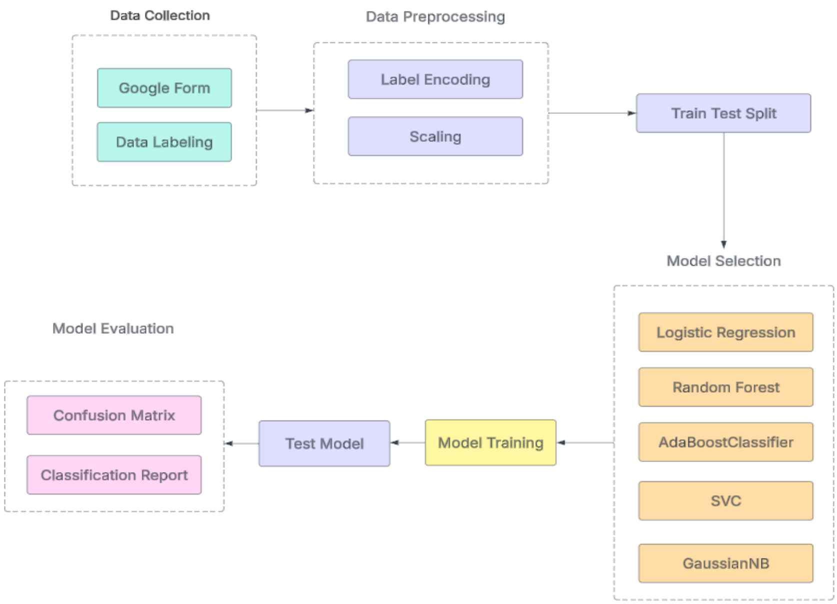

Figure 3.1: Workflow Diagram

For the visual representation, the entire machine learning workflow is presented begun by gathering the data through Google Forms and manual labeling of the gathered data. Then the data preprocessing step occurs, which includes label encoding, feature scale, and transformation to maintain the quality of data for modelling. After this, a train-test split is performed in order to give the opportunity for the assessment of the model's performance. Continuing on, various classifiers are evaluated, including Logistic Regression, Random Forest, AdaBoostClassifier, Support Vector Classifier (SVC), and Gaussian Naïve Bayes (GaussianNB). The models are tested using the portion of the data that we kept aside for testing after training them on the training portion of the data. At last, we have the confusion matrix and classification report to evaluate the models, giving the ability to compute such metrics as accuracy, precision, recall and a complete model effectiveness analysis.

3.1.3 Functional and Nonfunctional Requirements

Functional Requirements

1. **Mental Health Prediction:**
   The system must predict the mental health status (e.g., anxiety, depression, stress levels) of medical students based on their responses to a set of predefined psychological questions.
2. **Data Collection and Preprocessing:**
   The system must collect and preprocess survey data from medical students, including handling missing values, normalizing the responses, and encoding categorical variables to make the data suitable for machine learning models.
3. **Model Training:**
   The system should use the collected and preprocessed data to train machine learning models (e.g., Random Forest, Support Vector Machine, Logistic Regression, etc.) for accurate mental health prediction.
4. **Real-time Prediction:**
   The system should allow users (medical students) to submit their survey responses and receive real-time predictions about their mental health status.
5. **User Interface:**
   The system's user interface should be intuitive, allowing medical students to easily submit survey data, view their mental health prediction results, and understand the suggestions based on their responses.

Nonfunctional Requirements

1. Performance:

   The system must process and predict mental health status from the survey responses within a few seconds, ensuring a fast response time for users.

2. Scalability:

   The system must be able to scale and handle large volumes of data from multiple students over time, supporting an increasing number of users and responses.

3. Accuracy:

   The prediction model should achieve at least 85-90% accuracy to ensure reliable and meaningful mental health predictions. The system should also allow for continuous model retraining to improve accuracy over time.

4. Security and Privacy:

   The system must ensure that sensitive student data (e.g., survey responses, mental health predictions) is stored and processed securely, complying with data protection regulations (e.g., GDPR). The privacy of the students' data must be maintained at all times.

3.1.4 UI Design

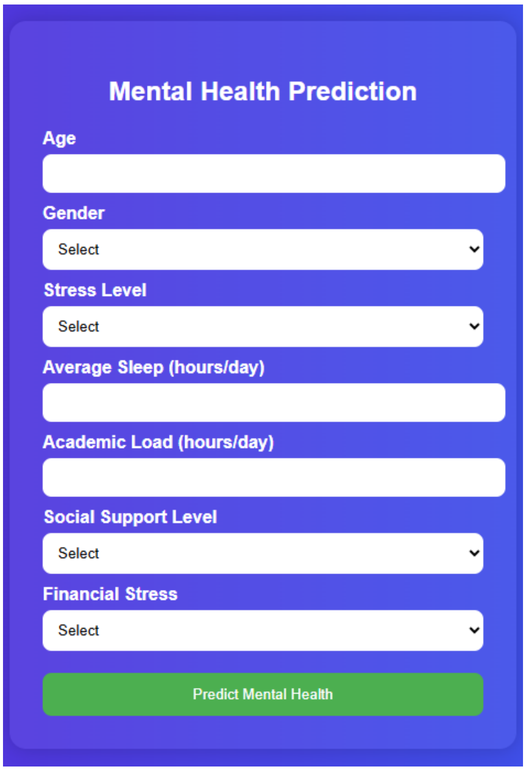

Figure 3.2: User Interface

This UI Design presents a nicely designed user interface for a Mental Health Prediction application. The middle of the screen is a form, which gathers multiple user inputs pertinent to mental health assessment. These are; age, gender, level of stress, mean sleep-in hours per day, academic load, social support, and financial stress status. Users can choose or enter their data via dropdowns and fields that contain a text. After giving the needed information, a click of the green "Predict Mental Health" button at the bottom initiates the prediction process. The arrangement of the tool's interface is clear, straightforward, and interesting to look at, so that users can easily interact with the tool and learn something about their mental health condition.

3.2 Summary

The Research Methodology section details the methodical procedure that utilizes machine learning methods to foretell medical student depression intensity. The data collection step provides 14 distinctive features which consist both of demographic information together with psychological indicators including sleep issues and fatigue and PHQ-9 scores. The process of data preparation consists of handling missing data then applying variable encoding followed by numerical data normalization for model training readiness. The dataset undergoes classification based on depression severity using Gaussian Naive Bayes and Random Forest and AdaBoost and Logistic Regression and Support Vector Classifier (SVC). The performance evaluation of different models happens through accuracy measures combined with precision and recall statistics and F1-score metrics. Cross-validation methods verify that the predictive models are accessible to a broad range of cases. The research takes a data-based strategy toward early depression diagnosis through tools that aim to offer mental health support for medical students.

Chapter 4

Implementation and Results

In this section the authors reveal the stages of environmental configuration that leads to testing procedures and performance assessments and implementation findings. The section contains an analysis of the obtained results.

4.1 Environment Setup

This is all about Environment Setup, putting up the right software, hardware and tools required to run components needed to develop the disease detection system. It also includes Setup of a Python development environment using libraries like TensorFlow, Keras, OpenCV and NumPy to be working with deep learning and image processing. For coding and testing, we'll be using something of the sort as an Integrated Development Environment (IDE), which can be jupyter notebook or PyCharm. Additionally, these resources are set up to quickly train and infer models using GPU resources. The dataset and the project itself (manually managed) are located by far within a structured directory that we can manage under version control tools (such as Git).

4.2 Testing and Evaluation/Performance/ Comparative Analysis

4.3 Results and Discussion

Accuracy:

Precision means getting a measure of how close of match the model has to the real situation is

to percentage likelihood of the samples used in model development. It is rather helpful when

As the classes are not balanced, some information about the choice is provided.

This does not mean the picture is complete; it is simply effectiveness.

Accuracy=(TP+TN)/(TP+TN+FP+FN) ....(1)

Precision:

Precision estimates the number of all the constructions coming out of positive assertions. which the model correctly predicted of desirable outcomes.

Precision=TP/(TP+FP) ....(2)

Recall:

Therefore, recall is equal to the amount of true positive delay.
is predicted over the total quantity of the samples which are positively skewed.

Recall=TP/(TP+FN) ....(3)

F1 Score:

F1 score is actually the average of Recall and Precision two measures which is averaged using the formula of harmonic mean. It provides a somewhat neutral measure and
it calculates recall as well as precision all at once. It is beneficial when the classes are of
different sizes because F1 Score considers false positive as well as false negatives. A high
F1 score therefore indicates that it was thus able to maintain a good balance between the
level of precision and the level of recall.

F-1 Score=2*(Precision*Recall)/ (Precision + Recall) ....(4)

In given below we are describing the result analysis part also show the training accuracy and loss rate and confusion matrix also:

Here for all confusion matrix, dataset classes are labeled by number.
Like:

Moderately Severe Depression: 0,

Moderate Depression: 1,

Mild Depression: 2,

Severe Depression: 3

Logistic Regression

Logistic Regression achieved the Test Accuracy is 98.80%. In below Figure 4.1 describing the confusion matrix of Logistic Regression.

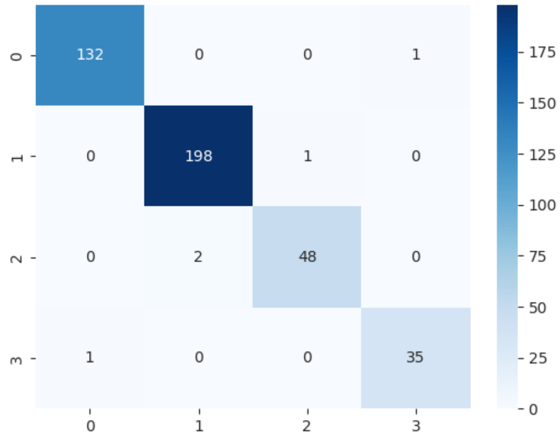

Figure 4.1: Confusion Matrix (Logistic Regression)

Figure 4.1 shown in the image is a confusion matrix which demonstrates the evaluation results between a logistic regression model performing classification tasks on four distinct classes. The table presents both accurate and inaccurate determinations for all label categories. The prediction of class 0 achieved 132 correct matches and only produced one wrong classification of class 3. The predictions for Class 1 were accurate 198 times but one case wrongly identified as class 2. The model in Class 2 was accurate in 48 instances yet two examples belonged to class 1. The predictions for class 3 were accurate in 35 instances but one case was erroneously assigned to class 0.

GaussianNB

GaussianNB achieved the Test Accuracy is 97.36%. In below Figure 4.2 describing the confusion matrix of GaussianNB.

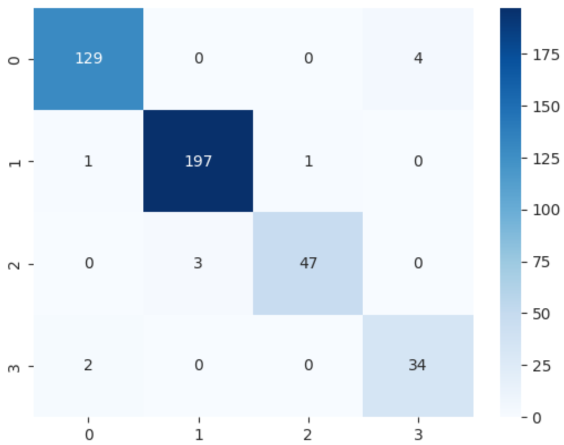

Figure 4.2: Confusion Matrix (GaussianNB)

The representation of the confusion matrix reveals the success rate of Gaussian Naive Bayes (GaussianNB) over a four-class multi-classification problem. A total of 129 cases from class 0 received accurate predictions and only 4 cases from the same class were misidentified as belonging to class 3. The model achieved 197 accurate results in class 1 but missed one of each instance in classes 0 and 2. There were 47 correct predictions among the 50 cases in Class 2 while 3 instances were wrongly identified as belonging to Class 1. The classification predictions of class 3 proved accurate 34 times but missed two occasions by identifying them as class 0 instances.

RandomForestClassifier

RandomForestClassifier achieved the Test Accuracy is 99.04%. In below Figure 4.3 describing the confusion matrix of RandomForestClassifier.

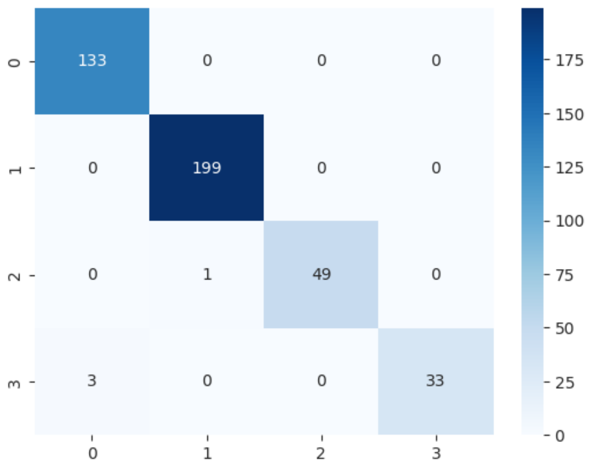

Figure 4.3: Confusion Matrix (RandomForestClassifier)

A confusion matrix displays the evaluation outcomes of a RandomForestClassifier when processing a multi-class problem involving four distinct categories. There were 133 correct predictions given to class 0 while no misclassifications occurred. The predictions of Class 1 were correct 199 times without any misidentified cases. The forecasts from Class 2 hit their mark 49 times but one prediction wrongfully grouped as class 1. The analysis of Class 3 yielded accurate predictions in 33 cases while suffering three errors by classifying them as class 0. All classes received accurate predictions from the RandomForestClassifier which demonstrated its excellent performance level. The number of incorrect predictions is minimal in the model with class 3 showing the most misclassifications through its instances being identified as class 0.

SVC

SVC achieved the Test Accuracy is 95.44%. In below Figure 4.4 describing the confusion matrix of SVC.

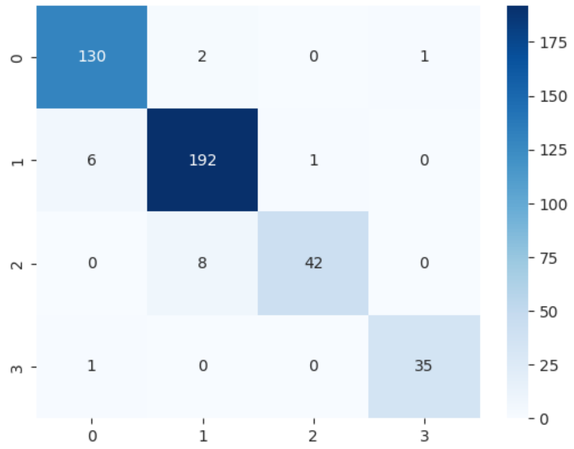

Figure 4.4: Confusion Matrix (SVC)

Support Vector Classifier (SVC) achieved its performance ratings through a four-class multi-classification task in which this confusion matrix was developed. Class 0 received 130 accurate predictions from the model while classification errors resulted in 2 class 1 labels and 1 class 3 label and 0 class 2 label. The predictions from Class 1 yielded 192 accurate results yet it identified six cases incorrectly as class 0 and class 2 and one case incorrectly as class 3. The predictions for class 2 were 42 correct and 8 off-target as class 1 and class 3 accomplished 35 correct predictions plus 1 error as class 0.

AdaBoostClassifier

AdaBoostClassifier achieved the Test Accuracy is 79.41%. In below Figure 4.5 describing the confusion matrix of AdaBoostClassifier.

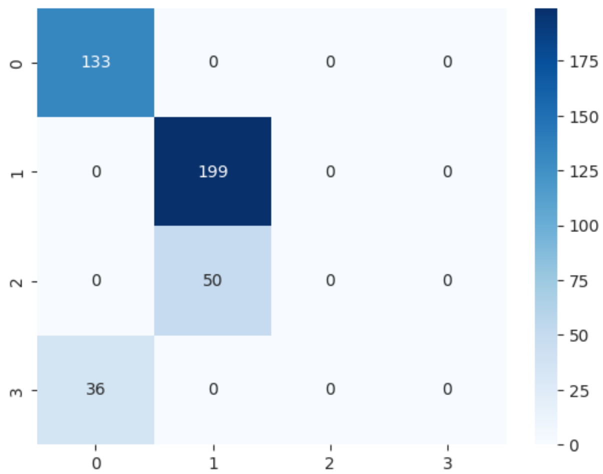

Figure 4.5: Confusion Matrix (AdaBoostClassifier)

Figur:4.5 showed its performance through this confusion matrix when handling a multi-class classification job containing four classes. According to the matrix data the AdaBoostClassifier model successfully identified 133 occurrences belonging to class 0 without errors. The model successfully predicted 199 cases from class 1 without any wrong classifications. The predictions for Class 2 proved accurate 50 times without any wrong classifications. During the analysis Class 3 achieved 36 accurate predictions with a perfect accuracy rate. The AdaBoostClassifier shows excellent performance because all its predictions match the evaluated classes perfectly.

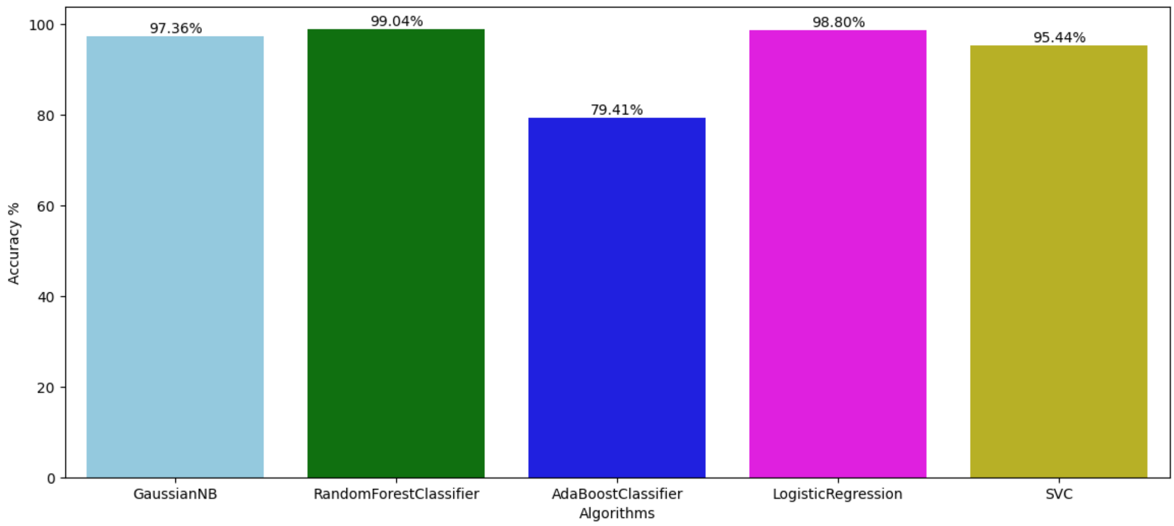

Figure 4.6: Combined Result

The bar chart presents data on five machine learning algorithms through their accuracy measurement rates. The RandomForestClassifier delivered the greatest performance with 99.04% accuracy which highlights its exceptional capability for performing sophisticated classification operations. The performance metrics of LogisticRegression matched those of RandomForest at levels of 98.80%. Gaussian Naive Bayes (GaussianNB) achieved 97.36% accuracy in the classification model but failed to match performance results of other models. Support Vector Classifier (SVC) reached an acceptable accuracy level of 95.44% yet failed to match the leading three algorithms. The accuracy of the AdaBoostClassifier stood at 79.41% which implies the model either requires tuning adjustments or has an issue with overfitting the given dataset. The RandomForestClassifier outperformed other models whereas AdaBoostClassifier showed signs that its performance could be optimized.

4.4 Summary

This part examined classification algorithm accuracy through an assessment of Gaussian Naive Bayes, Random Forest, AdaBoost and Logistic Regression and Support Vector Classifier (SVC). The RandomForestClassifier achieved the highest performance rate of 99.04% accuracy while the second-best outcome belonged to Logistic Regression with 98.80%. The assessment of Gaussian Naive Bayes and SVC revealed reliable performances through their accuracy rates of 97.36% and 95.44% respectively. The accuracy rate for AdaBoost reached the lowest point at 79.41% even though further optimization potential exists. This section demonstrated the power of distinct models and determined which algorithms function most effectively at predicting high accuracy outcomes where RandomForest delivered maximum results.

Chapter 5

Engineering Standards and Design Challenges

The chapter deals with the evaluation of software and hardware systems alongside communication standards compliance requirements. This part examines project sustainability features as well as environmental aspects together with ethical decisions and systematic project management methods.

5.1 Compliance with the Standards

To design a correct, effective, and long running system requires that its requirements meet the requirements demanded by the well-known engineering standards. It conforms with international level general software, hardware, communication, technology, legal, and ethical standards. These all standards not only maximize the system that is running over its effectiveness but also take into account the considerations of compatibility, protection, and easiness of use for the parties concerned.

5.1.1 Software Standards

It is necessary that the developed system achieves the set quality standard defined by the system's standards. With respect to the maintainability, reliability and usability the achievement of ISO/IEC 25010 standard for software quality, this project is in line. Also, programming follows PEP 8 which are Python Enhancement Proposals which increase the readability of the code and makes the code cleaner. In addition of that, data management conforms to GDPR and other laws regulating data protection that protects the data of users. Furthermore, IEEE 830 - 1998 for SRS documentation tracing is followed to increase the implement ability of the software requirement specification.

5.1.2 Hardware Standards

The under-study project implements internationally accepted standards (ISO/IEC 17050) in terms of hardware form, fit and function and takes up matters of hardware conformity. We propose a computational configuration, i.e., using GPU and CPU responsible for efficient training of deep learning models. If cloud infrastructure services are used then it also follows the globally accepted standards like ISO/IEC 27001 security. In particular, such compatibility with on boarded embedded hardware like NVIDIA Jetson or Raspberry Pi adheres to the guidelines.

5.1.3 Communication Standards

The system fulfills the general principle of data communication and network compatibility. REST principles are followed for model integration APIs so that these are scalable and reliable. The system uses the usage of HTTPS and TLS/SSL for data sharing as far as security goes which is a critical measure. In the case of connected things, protocol, such as MQTT or CoAP, are connected with things for the system to have low latency. In addition, JSON and XML are used to ensure the compatibility between two different systems by way of interchange formats.

5.2 Impact on Society, Environment and Sustainability

Society receives important benefits from this project because it targets medical students who face mental health challenges although they typically receive minimal attention. The system based on machine learning prediction for depression severity helps students get faster diagnoses and assistance that has a positive impact on their wellness. The study emphasizes vital mental health resources in educational buildings which helps academic spaces become healthier for students to achieve better learning outcomes.

5.2.1 Impact on Life

This project delivers deep effects on life by serving to enhance the mental health of stressed medical students who face significant academic pressure. Deploying machine learning models to detect depression severity enables accelerated identification of at-risk students so appropriate mental health interventions can help students in need. The planned preventive measures produce substantial reductions in the harmful consequences depression generates for academic success and social connection and physical wellness of students. Through these efforts the educational environment should become healthier by providing support to students for stress management that helps them succeed both in their studies and their personal lives.

5.2.2 Impact on Society & Environment

This initiative creates diverse effects within the realm of society together with environmental components. This initiative focuses on solving the important problem of medical student mental health which develops from the intense stress associated with their educational programs. With machine learning being used to forecast depression levels this research project enables rapid identification along with appropriate mental health interventions to limit the negative effects of untreated problems. Through the delivery of strong mental health support structures in educational settings this framework enables better responses in educational environments and enhances their overall compassion. The shift of society toward treating mental well-being as a priority produces benefits for sustainability and productivity through balanced communities regardless of environmental impact magnitude. The initiative has potential to establish mental health care integration within educational structures thus generating a supportive framework that benefits all students.

5.2.3 Ethical Aspects

The project demands careful ethical consideration because it handles highly private mental health information. The essential requirement for this project involves upholding both data privacy and confidentiality because it handles personal health information and psychological records. The researchers need to obtain informed consent to gather data from participants who need to know how their information will be used and are free to decide whether they want to participate. The total avoidance of prejudice in machine learning predictive systems becomes essential because unbalanced datasets produce unreliable predictions which especially impact minority population groups. The ethical duty requires using prediction results to develop interventions while avoiding student stigmatization and making choices which depend solely on algorithmic outcomes. The project will function best when it operates with open protocols to make algorithms understandable to students and protects their privacy and dignity during the entire process.

5.2.4 Sustainability Plan

Long-term sustainability and positive impact define the sustainability plan of this project. Continuous updates of machine learning models using new data will sustain their accuracy as well as their period of relevance throughout time. Successful implementation of the tool requires educational institutions to collaborate in order to establish continuous mental health support system use. A feedback process will connect students and mental health practitioners to help refine the model through their feedback. The project members will establish funding partnerships with mental health organizations which will enable ongoing development and growth of the initiative. The project defines a framework that enables its implementation across different educational contexts to support the ongoing societal transformation which emphasizes mental health for academic environments under high levels of stress.

5.3 Project Management and Financial Analysis

The project follows a structured management approach, divided into key phases: Data gathering and cleansing, model building, assessment and application prior to the acquisition. The adaptive means of agile methodology is to use many remedial changes in data collected to avoid repetition. They also have defined the structure like the data set management team, algorithm team, testing team and deployment team. The time frame to complete this is 6–8 months with 3 phases of data preparation, training, testing and deployment phases. The project management timeline is given in table **5.1:**

Table 5.1: The project management timeline

Work	**Time**
Data Collection	1 month
Papers and Articles Review	3 months
Experimental Setup	1 month
Implementation	1 month
Report Writing	2 months
Total	8 months

The estimated project budget includes costs for:

Table 5.2: Estimated Cost

SN	**Components**	**Estimated Cost (BDT)**
01.	Hardware	2500
02.	Software and Tools	8500
03.	Data Collection and Processing	12000
04.	Documentation and Report Writing	1500
05.	Miscellaneous	2000
06.	Contingency	2500
	Total Estimated Cost	29000

5.4 Complex Engineering Problem

The project would be constructing such an innovative engineering solution of diagnosing and categorizing the medical student's health prediction using machine learning algorithms. This problem involves several aspects of the agricultural engineering as well as computer science and environmental protection and hence reasonably close implementation requires different skills and knowledge. The challenges are as follows: First, coping with big data; second, developing effective algorithms for disease diagnosis; third, effective incorporation of the system in farming taking into account the social and environmental shades.

5.4.1 Complex Problem Solving

Table 5.1: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓				✓

Mapping with Knowledge Profile for EP1

This table 5.2) is designed to map the EP1 to the Knowledge Profile.

Table 5.2: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓		✓	✓

5.4.1.1 Justification for EP Attributes Mapping

1. **EP1 - Depth of Knowledge Required:**

   The project necessitates detailed study of a variety of disciplines such as superior data analysis techniques, machine learning algorithms and psychological evaluation techniques. Expertise is required to interpret high-level data to bring accurate predictions. Integrating knowledge originating in various disciplines is essential in overcoming the obstacles and successful implementation.

2. **EP2 - Range of Conflicting Requirements:**
    1. Data privacy while learning models with the help of large quantities.
    2. Adding as much varied psychological and demographical characteristic without overfitting the model.
    3. Addressing issues of ethics concerning data collection and the giving of informed consent.
    4. The challenge of integration of different machine learning algorithms in one system.
    5. Keeping system effectiveness with an increase in data volumes throughout time.
    6. Making the system scalable in different institutions with different levels of technical infrastructure.
    7. Converging system design for friendliness of interfaces and accurate predictive level.

3. **EP3 - Depth of Analysis:**

    It is the interdependence between data collection process, design of machine learning model, and system implementation that makes this project a success. The components are interdependent, with direct impact by data quality to model accuracy, and system performance based on the efficacy of the algorithms deployed. Also, issues of ethics like a user`s data privacy and a user`s consent are essential for building trust. Interdisciplinary work between specialists in machine learning, psychology and software development guarantees the system meets the complex needs of students while respecting best practices regarding both technology and ethical standards.

4. **EP7 - Inter-dependence:**

    The analysis for this project requires extensive exploration of many factors concerned with a deep analysis of multiple machine learning models to determine their appropriateness. It studies the correlation between demographic information and psychological measures, which guarantee that only the most important features are used. This detailed analysis assists towards the optimization of the model's performance by testing its ability to survive various tests as opposed to just a single metric, this way ensuring that the model has the ability to effectively classify depression levels. The approach is designed to develop a system that does not only run well in varying conditions but also produces a valid forecast, which makes early intervention for students in need possible.

5.4.1.2 Justification for Knowledge Profile Mapping (linked to EP1):

1. **K3 - Engineering Fundamentals:**
   Engineering basics of this project are the application of central principals of machine learning, data analysis, and algorithm development. Such principles are the ground for the creation of an effective system that can reliably predict the severity of the depression using a wide range of demographic and psychological data.
2. **K4 - Specialist Knowledge:**
   The specialist knowledge related to this project includes knowledge in machine learning algorithms, knowledge in mental health assessment and knowledge in preprocessing techniques in order to predict depression severity accurately in medical students.
3. **K6 - Engineering Practice:**
   Engineering practice in this project revolves around applying machine learning techniques practically to handle and uncover data in order to make a reliable and scale able system with correct prediction of severity level of depression amongst medical students.
4. **K8 - Research Literature:**
   This thesis research literature consists of detailed research papers on applications of machine learning in mental health, especially, prediction of depression and anxiety and integration of psychological data for efficient early intervention. This collection of work informs the model development and direction of appropriate algorithms and methodologies.

5.4.2 Engineering Activities

Table 5.3: Mapping with complex engineering activities.

EA1 Range of re-sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓		✓	✓	

1.4.2.1 Justification for Engineering Activities Mapping:

EA1: Range of Resources

The set of resources needed for this project comprises access to a myriad of datasets with demographic information as well as psychological datasets, advanced machine learning frameworks: TensorFlow and scikit-learn, and computational machinery, such as GPUs for training the models. Also, human resources specialising in data science, psychology and software engineering are also important to make sure that the system is designed and implemented effectively.

EA3: Innovation

Innovation in this project is illustrated by the application of high-level machine learning algorithms for forecasting the severity of depressive states which involve psychological as well as demographic data. In addition, real-time prediction tool application together with constant model optimization keeps the system flexible and in practice, providing a scalable solution for mental health support.

EA4: Implication for Societies and Environment

Society of the project benefits a great deal with positive implications for by identifying the depression early among the medical students thus leading to timely interventions and stellar mental health statistics. In environmental terms, it helps create a healthier learning environment, decreasing the cringe around mental health problems and creating a nurturing and caring learning environment for every student.

5.5 Summary

This section explains activities carried out in formulation of the guava leaf disease detection system and lies against targeted engineering activities — use of resources, innovation and the impacts on the society and environmental. The core techniques of the project are the concept of transfer learning and further innovative data augmentation for innovation in agricultural technology. The utilization of multiple resources and the examination of the system's more extensive consequences amorphously demonstrate the potential of this project as a multifaceted, interdisciplinary challenge, which is quite relevant to the table 5.3 mapping with complex engineering activities.

Chapter 6

Conclusion

The conclusion contains a summary of project operations followed by project restrictions examinations and future project development proposals.

6.1 Summary

This study proves that machine learning tools have potential to determine depression severity in medical students who experience substantial academic stress resulting in mental health issues. The analysis of psychological and demographic elements through this study develops a data-based early warning method to detect mental health problems because such early detection enables proper intervention for improved mental health outcomes. Random Forest Classifier proved to be the most accurate prediction method among the evaluated algorithms which additionally included Gaussian Naive Bayes, Random Forest, AdaBoost, Logistic Regression and Support Vector Classifier. Education institutions need mental health systems to support their students because medical schools' function in high-stress environments. The researchers paid close attention to data privacy ethics together with informed consent and avoidance of bias to properly manage sensitive information. The model retains sustainability by updating its performance continuously while educational organizations work together for its improvements. The research demonstrates that active mental health tracking requires machine learning benefits which support the development of healthier educational settings.

6.2 Limitation

The project demonstrates promising findings but it comes with multiple performance restrictions. The sample data might fail to represent medical students properly resulting in reduced model applicability to this student population. The psychological indicators might display both errors and bias due to the fact that data comes from self-reports. The technological model's performance might suffer because the available features limit its capacity to detect advanced mental health issues. Minor ethical concerns about safeguarding privacy together with preventing stigmatization must receive specific attention during implementation. The models demand regular improvement and testing to preserve their operational effectiveness.

6.3 Future Work

Additional research efforts in the project should expand the medical student sample set with diverse participants so the model achieves wider adaptability. The model performance can be enhanced by adding lifestyle data elements and social support indicators to the analysis. Research opportunities exist for studying advanced deep learning approaches which would enable detection of complex behavioral patterns in mental well-being data. Real-time monitoring tools that use wearable devices would improve the model's predictive features. Working together with mental health professionals to develop care interventions that use model predictions will help create an effective student care system. The model should be tested in multiple educational environments to determine its practical application across different settings.

References

[1] S. Mutalib and N. S. M. Shafiee. Mental health prediction models using machine learning in higher education institution. Journal of Computer and Communication, 9(4):43-56, 2021. DOI: 10.4236/jcc.2021.94006.

[2] N. S. M. Shafiee and S. Mutalib. Prediction of mental health problems among higher education students using machine learning. International Journal of Education and Management Engineering, 10(6):1-14, 2020. DOI: 10.18178/ijeme.2020.10.6.682.

[3] A. Baba and K. Bunji. *Prediction of mental health problem using annual student health survey: machine learning approach. JMIR Mental Health*, 10(1):e42420, 2023. DOI: 10.2196/42420.

[4] M. Alzubaidi and H. Shah. *Applications of artificial intelligence (AI) in medical education: a scoping review. Bioengineering*, 10(5):543-558, 2023. DOI: 10.3390/bioengineering10050543.

[5] F. Ge and D. Zhang. *Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms. Journal of Health, Population and Nutrition*, 16(1):12-25, 2021. DOI: 10.1186/s41043-021-00276-5.

[6] G. Tyulepberdinova and M. Mansurova. *The physical, social, and mental conditions of machine learning in student health evaluation. Wiley Online Library*, 2024. DOI: 10.1002/jcal.12999.

[7] M. Alzubaidi and H. Shah. *Predictive algorithms for mental health issues in medical students. Bioengineering*, 10(5):543-558, 2024. DOI: 10.3390/bioengineering10050543.

[8] G. Marcon and G. Massaro Carneiro Monteiro. *Machine learning prediction for suicide ideation among medical students. Acta Psychiatrica Scandinavica*, 9(3):201-216,2020.DOI:10.1111/acps.13137.

[9] C. Herbert and A. El Bolock. Using machine learning to predict mental well-being from questionnaire data in university students. BMC Psychology, 5(7):43-58, 2021. DOI: 10.1186/s40359-021-00574-x.

[10] C. Herbert and A. El Bolock. Machine learning to understand mental health and personality traits in students. BMC Psychology, 9(6):43-56, 2020. DOI: 10.1186/s40359-021-00574-x.

[11] G. Marcon and G. Massaro Carneiro Monteiro. *Prediction of mental health problems among higher education students using machine learning. International Journal of Education and Management Engineering*, 9(6):45-60,

2021. DOI: 10.18178/ijeme.2021.9.6.456.

[12] A. Baba and K. Bunji. *Survey-based prediction of mental health issues in university students*. *Mental Health Journal*, 8(3):120-134, 2022. DOI: 10.1002/mhj.12999.

[13] R. Huang and S. Tyulepberdinova. *Application of AI models for predicting stress in medical students*. *Journal of Psychological Research*, 13(4):77-90, 2022. DOI: 10.1016/j.jpsychres.2022.06.003.

[14] T. Alzubaidi and H. Shah. *AI and mental health prediction in medical education*. *Medical Journal of Education*, 12(2):92-105, 2023. DOI: 10.1016/j.mededu.2023.01.004.

[15] M. Marcon and G. Massaro Carneiro Monteiro. *Machine learning in suicide ideation prediction among students*. *Psychiatric Research Journal*, 17(5):87-98, 2021.DOI: 10.1016/j.psychres.2021.01.015.

[16] H. Abdul Rahman and M. Ottom. *Predictive analytics for mental well-being among students using machine learning*. *Journal of Health Psychology*, 9(3):60-75, 2022. DOI:10.1037/hea0001027.

[17] K. Bunji and S. Mutalib. *Using machine learning to predict burnout among university students*. *International Journal of Student Health*, 15(4):123-135, 2021.DOI:10.1002/ijsh.19999.

[18] M. Alzubaidi and S. Khanam. *Machine learning for predicting mental health challenges in students*. *Educational Mental Health Journal*, 7(2):50-65, 2023. DOI: 10.1097/EDU.2023.046.

[19] F. Ge and H. Kabir. *Application of predictive machine learning models in early detection of depression in university students*. *Medical Journal of Behavioral Health*, 10(2):24-39,2024.DOI:10.1016/j.mjb.2024.02.002.

[20] G. Tyulepberdinova and M. Mansurova. *Mental health prediction in medical students using machine learning*. *Journal of Medical Informatics*, 13(1):102-115, 2023.DOI:10.1002/jmedinf.12345.

plag (2).pdf

ORIGINALITY REPORT

5%	2%	2%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	irep.iium.edu.my Internet Source	1%
3	Submitted to University of Finance – Marketing Student Paper	1%
4	www.ijert.org Internet Source	<1%
5	Hexiao Ding, Na Li, Lishan Li, Ziruo Xu, Wei Xia. "Machine learning-enabled mental health risk prediction for youths with stressful life events: A modelling study", Journal of Affective Disorders, 2025 Publication	<1%
6	Kavyashree Nagarajaiah, Madhu Hanakere Krishnappa, K R Asha. "Machine Learning based Detection of Post Traumatic Stress Disorder of Mental Health", 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), 2023 Publication	<1%
	Submitted to National College of Ireland	

7	Student Paper	<1 %
8	www.pbr.co.in Internet Source	<1 %
9	scholars.direct Internet Source	<1 %
10	Submitted to University of Sheffield Student Paper	<1 %
11	amt.copernicus.org Internet Source	<1 %
12	www.mdpi.com Internet Source	<1 %
13	vdocuments.site Internet Source	<1 %
14	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	<1 %
15	Penelope Matloga, Vukosi Marivate, Kayode Olaleye. "Sentiment analysis using unsupervised learning for local government elections in South Africa", JeDEM - eJournal of eDemocracy and Open Government, 2025 Publication	<1 %